Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2010, Article ID 719197, 11 pages
doi:10.1155/2010/719197

Research Article

A Stereo Crosstalk Cancellation System Based on the Common-Acoustical Pole/Zero Model

Lin Wang,[1,2] Fuliang Yin,[1] and Zhe Chen[1]

[1] *School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116023, China*
[2] *Institute for Microstructural Sciences, National Research Council Canada, Ottawa, ON, Canada K1A 0R6*

Correspondence should be addressed to Lin Wang, wanglin_2k@sina.com

Received 8 January 2010; Revised 21 June 2010; Accepted 7 August 2010

Academic Editor: Augusto Sarti

Crosstalk cancellation plays an important role in displaying binaural signals with loudspeakers. It aims to reproduce binaural signals at a listener's ears via inverting acoustic transfer paths. The crosstalk cancellation filter should be updated in real time according to the head position. This demands high computational efficiency for a crosstalk cancellation algorithm. To reduce the computational cost, this paper proposes a stereo crosstalk cancellation system based on common-acoustical pole/zero (CAPZ) models. Because CAPZ models share one set of common poles and process their zeros individually, the computational complexity of crosstalk cancellation is cut down dramatically. In the proposed method, the acoustic transfer paths from loudspeakers to ears are approximated with CAPZ models, then the crosstalk cancellation filter is designed based on the CAPZ transfer functions. Simulation results demonstrate that, compared to conventional methods, the proposed method can reduce computational cost with comparable crosstalk cancellation performance.

1. Introduction

A 3D audio system can be used to position sounds around a listener so that the sounds are perceived to come from arbitrary points in space [1, 2]. This is not possible with classical stereo systems. Thus, 3D audio has the potential of increasing the sense of realism in music or movies. It can be of great benefit in virtual reality, augmented reality, remote video conference, or home entertainment. A 3D audio technique achieves virtual sound perception by synthesizing a pair of binaural signals from a monaural source signal with the provided 3D acoustic information: the distance and direction of the sound source with respect to the listener. Specifically, the sense of direction can be rendered by using head-related acoustic information, such as head-related transfer functions (HRTFs) which can be obtained by either experimental or theoretical means [3, 4]. To deliver binaural signals, the simplest way is through headphones. However, in many applications, for example, home entertainment environment, teleconferencing, and so forth, many listeners prefer not to wear headphones. If loudspeakers are used, the delivery of these binaural signals to the listener's ears is not straightforward. Each ear receives a so-called crosstalk component, moreover, the direct signals are distorted by room reverberation. To overcome the above problems, an inverse filter is required before playing binaural signals through loudspeakers.

The concept of crosstalk cancellation and equalization was introduced by Atal and schroeder [5] and Bauer [6] in the early 1960s. Many sophisticated crosstalk cancellation algorithms have been presented since then, using two or more loudspeakers for rendering binaural signals. Crosstalk cancellation can be realized directly or adaptively. Supposing that the acoustical transfer paths from loudspeakers to ears are known, the direct implementation method calculates the crosstalk cancellation filter by directly inverting the acoustical transfer functions [7, 8]. Generally a head-tracking scheme, which can tell the head position precisely, is employed to work together with the direct estimation method. The direct estimation method can be implemented in the time or frequency domain. Time-domain algorithms are generally computationally consuming, while frequency-domain algorithms have lower complexity. On the other hand, time-domain algorithms perform better than

frequency-domain ones with the same crosstalk cancellation filter length. For example, a frequency-domain method such as the fast deconvolution method [7], which has been shown to be very useful and easy to use in several practical cases, can suffer from a circular convolution effect when the inverse filters are not long enough compared to the duration of the acoustic path response. In an adaptive implementation method, the crosstalk cancellation filter is calculated adaptively with the feedback signals received by miniature microphones placed in human ears [9]. Several adaptive crosstalk cancellation methods typically employ some variation of LMS or RLS algorithms [10–13]. The LMS algorithm, which is known for its simplicity and robustness, has been used widely, but its convergence speed is slow. The RLS algorithm may accelerate the convergence, but the large computation load is a side effect. Although many algorithms have been proposed, the adaptive implementation method remains academic research rather than a real solution. The reason is that people who do not want to use headphones would probably not like to use a pair of microphones in the ears to optimize loudspeaker reproduction either.

One key limitation of a crosstalk cancellation system arises from the fact that any listener movement which exceeds 75–100 mm may completely destroy the desired spatial effect [14, 15]. This problem can be resolved by tracking the listener's head in 3D space. The head position is captured by a magnetic or camera-based tracker, then the HRTF filters and the crosstalk canceller based on the location of the listener are updated in real time [16]. Although head-tracking systems can be employed, measures should still be taken to increase the robustness of the crosstalk cancellation system. It has been shown that the robust solution to this virtual sound system could be obtained by placing the loudspeakers in an appropriate way to ensure that the acoustic transmission path or transfer function matrix is well conditioned [17–19]. Robust crosstalk cancellation methods with multiple loudspeakers have been proposed [8, 20, 21]. Another approach adds robustness of a crosstalk canceller by exploring the statistical knowledge of acoustic transfer functions [22].

This paper focuses on the crosstalk cancellation problem for a stereo loudspeaker system. Least-squares methods are popular in designing a crosstalk cancellation system; however, the required large computation is always a challenge. To reduce the computational cost, this paper proposes a novel crosstalk cancellation system based on common-acoustical pole/zero (CAPZ) models, which outperforms conventional all-zero or pole/zero models in computational efficiency [23, 24]. The acoustic paths from loudspeakers to ears are approximated with CAPZ models, then the crosstalk cancellation filters are designed based on the CAPZ transfer functions. Compared with conventional least-squares methods, the proposed method can reduce the computation cost greatly. The paper is organized as follows. Conventional crosstalk cancellation methods are introduced in Section 2. Then the proposed crosstalk cancellation method based on the CAPZ model is described in detail in Section 3. The performance of the proposed method is evaluated in Section 4. Finally, conclusions are drawn in Section 5.

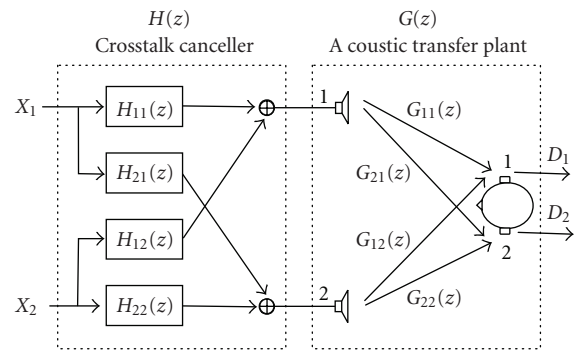


Figure 1: Block diagram of the direct crosstalk cancellation system for stereo loudspeakers.

2. Conventional Crosstalk Canceller

It is common to use two loudspeakers in a stereo system. A block diagram of the direct implementation of crosstalk cancellation is illustrated in Figure 1 for a stereo loudspeaker system. The input binaural signals from left and right channels are given in vector form $X(z) = [X_1(z), X_2(z)]^T$, and the signals received by two ears are denoted as $D(z) = [D_1(z), D_2(z)]^T$. (Here signals are expressed in the Z domain.) The objective of crosstalk cancellation is to perfectly reproduce the binaural signals at the listener's eardrums, that is, $D(z) = z^{-d}X(z)$, where z^{-d} is the delay term, via inverting the acoustic path $G(z)$ with the crosstalk cancellation filter $H(z)$. Generally, the loudspeaker response should also be inverted when designing the crosstalk canceller; however, this part can be implemented separately and thus is not considered in this paper for the convenience of analysis. $G(z)$ and $H(z)$ are, respectively, denoted in matrix forms as

$$G(z) = \begin{bmatrix} G_{11}(z) & G_{12}(z) \\ G_{21}(z) & G_{22}(z) \end{bmatrix}, \quad H(z) = \begin{bmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{bmatrix}, \tag{1}$$

where $G_{ij}(z)$, $i, j = 1, 2$, is the acoustic transfer function from the jth loudspeaker to the ith ear, and $H_{ij}(z)$, $i, j = 1, 2$, is the crosstalk cancellation filter from X_j to the ith loudspeaker.

To ensure crosstalk cancellation, the global transfer function from binaural signals to ears should be

$$D(z) = G(z)H(z)X(z) = z^{-d}X(z), \tag{2}$$

thus

$$G(z)H(z) = z^{-d}I\ , \tag{3}$$

$$H(z) = z^{-d}G^{-1}(z), \tag{4}$$

where I is the identity matrix. The delay term z^{-d} is necessary to guarantee that $H(z)$ is physical realizable (causal). However, a perfect reproduction is impossible because $G(z)$ is generally nonminimum-phase, in which case a least-squares algorithm is employed to approximate the optimal inverse filter $G^{-1}(z)$. The time-domain least-squares algorithm is given below.

Suppose that $g_{ij} = [g_{ij,0}, \ldots, g_{ij,L_g-1}]^T$, the time-domain impulse response of $G_{ij}(z)$, is a vector of length L_g, and $h_{ij} = [h_{ij,0}, \ldots, h_{ij,L_h-1}]^T$, the time-domain impulse response of $H_{ij}(z)$, is a vector of length L_h. Rewriting (3) in a time-domain form, we get

$$\begin{bmatrix} \widetilde{G}_{11} & \widetilde{G}_{12} \\ \widetilde{G}_{21} & \widetilde{G}_{22} \end{bmatrix} \cdot \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} u_d & O \\ O & u_d \end{bmatrix} \tag{5}$$

or in a suppressed form

$$GH = U, \tag{6}$$

where $\widetilde{G}_{ij}$, a component of G, is

$$\widetilde{G}_{ij} = \begin{bmatrix} g_{ij,0} & \cdots & g_{ij,L_g-1} & 0 & \cdots & 0 \\ 0 & g_{ij,0} & \cdots & g_{ij,L_g-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & g_{ij,0} & \cdots & g_{ij,L_g-1} \end{bmatrix}^T. \tag{7}$$

$\widetilde{G}_{ij}$ is a convolution matrix of size $L_1 \times L_h$ by cascading the vector g_{ij}, $L_1 = L_h + L_g - 1$,

$$u_d = [0, \ldots, 0, 1, 0, \ldots, 0]^T \tag{8}$$

is a vector of length L_1 whose dth component equals 1, and O is a vector of length L_1 containing only zeros.

The least-squares solution to (6) is

$$H_{\mathrm{LS}} = G^+ U, \tag{9}$$

where G^+ is the pseudoinverse of G, and G^+ is given by

$$G^+ = \left(G^T G + \beta I\right)^{-1} G^T, \tag{10}$$

where β is a regularization parameter to increase the robustness of the inversion [25].

The crosstalk cancellation filter is obtained by (9), with its filter length

$$L_{h1} = L_h. \tag{11}$$

The acoustic path matrix G is dependent on the head position. When the head moves, it is required to update G and calculate H in real time. The computation load becomes heavy when the size of G is large.

In [26], a single-filter structure for a stereo loudspeaker system is proposed to calculate the inverse of G, which needs less computation. It is given as follows.

From (4), we can get

$$\begin{aligned} H(z) &= z^{-d} G^{-1}(z) \\ &= \frac{z^{-d} \begin{bmatrix} G_{22}(z) & -G_{12}(z) \\ -G_{21}(z) & G_{11}(z) \end{bmatrix}}{G_{11}(z)G_{22}(z) - G_{12}(z)G_{21}(z)}. \end{aligned} \tag{12}$$

Let

$$Q(z) = G_{11}(z)G_{22}(z) - G_{12}(z)G_{21}(z), \tag{13}$$

$$T(z) = \frac{z^{-d}}{Q(z)}, \tag{14}$$

then the problem of inverting $G(z)$ is converted to

$$Q(z)T(z) = z^{-d} I. \tag{15}$$

Suppose that $q = [q_0, \ldots, q_{L_q-1}]^T$, the time-domain response of $Q(z)$, is a vector of length L_q, and $L_q = 2L_g - 1$; $t = [t_0, \ldots, t_{L_t-1}]^T$, the time-domain response of $T(z)$, is a vector of length L_t. Rewriting (15) in a time-domain form, we get

$$Qt = u_d, \tag{16}$$

where

$$Q = \begin{bmatrix} q_0 & \cdots & q_{L_q-1} & 0 & \cdots & 0 \\ 0 & q_0 & \cdots & q_{L_q-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & q_0 & \cdots & q_{L_q-1} \end{bmatrix}^T \tag{17}$$

is a convolution matrix of size $L_2 \times L_t$ by cascading of the vector q; $L_2 = L_t + L_q - 1$.

The least-squares solution to (16) is

$$t_{\mathrm{LS}} = Q^+ u_d, \tag{18}$$

where Q^+ is the pseudoinverse of Q, and Q^+ is given by

$$Q^+ = \left(Q^T Q + \beta I\right)^{-1} Q^T. \tag{19}$$

The crosstalk cancellation filter is obtained from (12) and (18), with its filter length

$$L_{h2} = L_t + L_g - 1. \tag{20}$$

Combining $G(z)$ and $H(z)$, we get the global transfer function

$$\begin{aligned} F(z) &= G(z) \cdot H(z) \\ &= T(z) \cdot \begin{bmatrix} G_{11}(z) & G_{12}(z) \\ G_{21}(z) & G_{22}(z) \end{bmatrix} \cdot \begin{bmatrix} G_{22}(z) & -G_{12}(z) \\ -G_{21}(z) & G_{11}(z) \end{bmatrix} \\ &= T(z) \\ &\quad \cdot \begin{bmatrix} G_{11}(z)G_{22}(z) & 0 \\ -G_{12}(z)G_{21}(z) & \\ 0 & G_{11}(z)G_{22}(z) \\ & -G_{12}(z)G_{21}(z) \end{bmatrix}. \end{aligned} \tag{21}$$

The off-diagonal items of (21) are always zeros regardless the value of $T(z)$. This implies that the crosstalk is almost fully suppressed. However, due to the filtering effect by the diagonal items in (21), distortion will be introduced when reproducing the target signals. This is the inherent disadvantage of the single-filter structure method.

3. Crosstalk Cancellation System Based on CAPZ Models

The acoustic transfer function is usually an all-zero model, whose coefficients are its impulse response. However, when the duration of the impulse response is long, it requires a large number of parameters to represent the transfer function [27]. This results in large computation in binaural synthesis and crosstalk cancellation. Pole/zero models may decrease the computational load, but their poles and zeros both change when the acoustic transfer function varies, leading to inconvenience for acoustic path inversion. To reduce the computational cost, this paper attempts to approximate the acoustic transfer function with common-acoustical pole/zero (CAPZ) models, then design a crosstalk cancellation system based on it.

3.1. CAPZ Modeling of Acoustic Transfer Functions. Haneda proposed the concept of common-acoustical pole/zero (CAPZ) models, and modeled room transfer functions and head-related transfer functions with good results [23, 24]. He believed that an HRTF contains a resonance system of ear canal whose resonance frequencies and Q factors are independent of source directions. Based on this, the HRTF can be efficiently modeled by using poles that are independent of source directions, with zeros that are dependent on source directions. The poles represent the resonance frequencies and Q factors. The model is called common-acoustical pole/zero model. CAPZ models share one set of poles and process their own zeros individually. This obviously reduces the amount of parameters with respect to conventional pole/zero models, and also cut down computation.

When an acoustic transfer function $H_i(z)$ is approximated with a CAPZ model, it is expressed as

$$\hat{H}_i(z) = \frac{B_i(\mathrm{z})}{A(z)} = \frac{\sum_{n=0}^{N_q} b_{n,i} z^{-n}}{1 + \sum_{n=1}^{N_p} a_n z^{-n}}, \tag{22}$$

where N_p and N_q are the numbers of the poles and zeros, $a = [1, a_1, \ldots, a_{N_p}]^T$ and $b_i = [b_{1,i}, \ldots, b_{N_q,i}]^T$ are the pole and zero coefficient vectors, respectively. The CAPZ parameters may be estimated with a least-squares method [23, 24] or a state-space method [28]. The least-squares method is simply given below.

Suppose a set of K transfer functions, the total modeling error is defined as

$$\begin{aligned} J &= \sum_{i=1}^{K} \sum_{n=0}^{N-1} |e_i(n)|^2 \\ &= \sum_{i=1}^{K} \sum_{n=0}^{N-1} \left| h_i(n) + \sum_{j=1}^{N_p} a_j h_i(n-j) - \sum_{j=0}^{N_q} b_{j,i} \delta(n) \right|^2, \end{aligned} \tag{23}$$

where N is the length of $e(n)$ and $h_i(n)$ is the impulse response of $H_i(z)$.

To find the pole coefficients vector a and the zero coefficients vector b_i, $i = 1, \ldots, K$, we minimize the error J and obtain that

$$\begin{aligned} \begin{bmatrix} I & H_{o,1} \\ 0 & H_1 \end{bmatrix} \begin{bmatrix} b_1 \\ -a \end{bmatrix} &= \begin{bmatrix} r_{o,1} \\ r_1 \end{bmatrix}, \\ &\vdots \\ \begin{bmatrix} I & H_{o,K} \\ 0 & H_K \end{bmatrix} \begin{bmatrix} b_K \\ -a \end{bmatrix} &= \begin{bmatrix} r_{o,K} \\ r_K \end{bmatrix}, \end{aligned} \tag{24}$$

where I is the identity matrix, vector $r_{o,i} = [h_i(0), \ldots, h_i(N_q)]^T$, $r_i = [h_i(N_q + 1), \ldots, h_i(N-1)]^T$, $i = 1, \ldots, K$; $H_{o,i}$ and H_i are both convolution matrices by cascading the impulse response $h_i(n)$, that is,

$$H_{o,i} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ h_i(0) & 0 & \cdots & 0 \\ h_i(1) & h_i(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_i(N_q - 1) & h_i(N_q - 2) & \vdots & h_i(N_q - N_p) \end{bmatrix}_{(N_q - 1) \times N_p}, \tag{25}$$

$$H_i = \begin{bmatrix} h_i(N_q) & \cdots & h_i(N_q - N_p + 1) \\ \vdots & \ddots & \vdots \\ h_i(N-2) & \cdots & h_i(N - 1 - N_p) \end{bmatrix}_{(N - 1 - N_q) \times N_p}. \tag{26}$$

From (24), a and b_i can be obtained by

$$\begin{aligned} a &= -\left(\tilde{H}^T \tilde{H}\right)^{-1} \tilde{H}^T R, \\ b_i &= H_{o,i} a + r_{o,i}, \qquad i = 1, \ldots, K, \end{aligned} \tag{27}$$

where vector $R = [r_1, \ldots, r_K]^T$ and matrix $\tilde{H} = [H_1, \ldots, H_K]^T$.

It is useful to specify the selection of the number of poles and zeros, N_p and N_q. The more poles and zeros used, the better approximation result may be obtained. On the other hand, more parameters require higher computation. Thus a trade-off should be considered. Generally, in the least-squares method, the number of parameters can be determined empirically [24]; or in the state-space method, it is determined based on the singular-value decomposition result [28].

3.2. Crosstalk Cancellation Based on the CAPZ Model. Supposing that acoustic transfer path G is known, the CAPZ

parameters are estimated. The CAPZ models from the loudspeakers to the ears are

$$\begin{aligned} G_{11}(z) &= \frac{B_{11}(z)}{A(z)}z^{-d_{11}}, \\ G_{12}(z) &= \frac{B_{12}(z)}{A(z)}z^{-d_{12}}, \\ G_{21}(z) &= \frac{B_{21}(z)}{A(z)}z^{-d_{21}}, \\ G_{22}(z) &= \frac{B_{22}(z)}{A(z)}z^{-d_{22}}, \end{aligned} \tag{28}$$

where d_{11}, d_{12}, d_{21}, and d_{22} are the transmission delays from the loudspeakers to the ears.

Substituting (28) into (4), we get

$$\begin{aligned} &H(z) \\ &= z^{-d}G^{-1}(z) \\ &= \frac{z^{-d}\begin{bmatrix} G_{22}(z) & -G_{12}(z) \\ -G_{21}(z) & G_{11}(z) \end{bmatrix}}{G_{11}(z)G_{22}(z) - G_{12}(z)G_{21}(z)} \\ &= z^{-d}/\left(\left(\frac{B_{11}(z)B_{22}(z)}{A^2(z)}\right)z^{-(d_{11}+d_{22})}\right. \\ &\qquad \left.-\left(\frac{B_{12}(\mathrm{z})B_{21}(\mathrm{z})}{A^2(z)}\right)z^{-(d_{12}+d_{21})}\right) \\ &\quad \times \begin{bmatrix} \left(\frac{B_{22}(z)}{A(z)}\right)z^{-d_{22}} & \left(-\frac{B_{12}(z)}{A(z)}\right)z^{-d_{12}} \\ \left(-\frac{B_{21}(z)}{A(z)}\right)z^{-d_{21}} & \left(\frac{B_{11}(z)}{A(z)}\right)z^{-d_{11}} \end{bmatrix} \\ &= \frac{z^{-d}}{B_{11}(z)B_{22}(z)z^{-(d_{11}+d_{22})} - B_{12}(z)B_{21}(z)z^{-(d_{12}+d_{21})}} \\ &\quad \times \begin{bmatrix} B_{22}(z)A(z)z^{-d_{22}} & -B_{12}(z)A(z)z^{-d_{12}} \\ -B_{21}(z)A(z)z^{-d_{21}} & B_{11}(z)A(z)z^{-d_{11}} \end{bmatrix}. \end{aligned} \tag{29}$$

Without loss of generality, assume $d_{11} + d_{22} < d_{12} + d_{21}$, and let $\Delta = (d_{11} + d_{22}) - (d_{12} + d_{21})$. Substituting Δ into (29), we get

$$\begin{aligned} H(z) &= \frac{z^{-(d-d_{11}-d_{22})}}{B_{11}(z)B_{22}(z) - B_{12}(z)B_{21}(z)z^{-\Delta}} \\ &\quad \times \begin{bmatrix} B_{22}(z)A(z)z^{-d_{22}} & -B_{12}(z)A(z)z^{-d_{12}} \\ -B_{21}(z)A(z)z^{-d_{21}} & B_{22}(z)A(z)z^{-d_{11}} \end{bmatrix} \\ &= \frac{z^{-\delta}}{B(z)}\begin{bmatrix} B_{22}(z)A(z)z^{-d_{22}} & -B_{12}(z)A(z)z^{-d_{12}} \\ -B_{21}(z)A(z)z^{-d_{21}} & B_{22}(z)A(z)z^{-d_{11}} \end{bmatrix} \\ &= C(z)\begin{bmatrix} B_{22}(z)A(z)z^{-d_{22}} & -B_{12}(z)A(z)z^{-d_{12}} \\ -B_{21}(z)A(z)z^{-d_{21}} & B_{11}(z)A(z)z^{-d_{11}} \end{bmatrix}, \end{aligned} \tag{30}$$

where $B(z) = B_{11}(z)B_{22}(z) - B_{12}(z)B_{21}(z)z^{-\Delta}$, $C(z) = z^{-\delta}/B(z)$, and $\delta = d - (d_{11} + d_{22})$ is the delay.

Thus the problem of inverting $G(z)$ is converted to

$$B(z)C(z) = z^{-\delta}I. \tag{31}$$

Suppose that $b = [b_0, \ldots, b_{L_b-1}]^T$, the time-domain impulse response of $B(z)$, is a vector of length L_b, and $L_b = 2(N_q + 1) + \Delta - 1$; $c = [c_0, \ldots, c_{L_c-1}]^T$, the time-domain impulse response of $C(z)$, is a vector of length L_c. Rewriting (31) in a time-domain form, we get

$$Bc = u_\delta, \tag{32}$$

where B is a convolution matrix of size $L_3 \times L_c$ by cascading the vector b, and $L_3 = L_b + L_c - 1$,

$$\begin{aligned} B &= \begin{bmatrix} b_0 & \ldots & b_{L_b-1} & 0 & \ldots & 0 \\ 0 & b_0 & \ldots & b_{L_b-1} & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & b_0 & \ldots & b_{L_b-1} \end{bmatrix}^T, \\ u_\delta &= [0, \ldots, 0, 1, 0, \ldots, 0]^T \end{aligned} \tag{33}$$

is a vector of length L_3 whose δth component equas 1.

Since $B(z)$ is generally nonminimum-phase, the least-squares solution to (32) is

$$c_{\mathrm{LS}} = B^+ u_\delta, \tag{34}$$

where B^+ is the pseudoinverse of B, and B^+ is given by

$$B^+ = \left(B^T B + \beta I\right)^{-1} B^T, \tag{35}$$

where β is the regularization parameter.

Finally, the crosstalk canceller is obtained by (30) and (34), with its filter length

$$\begin{aligned} L_{h3} &= L_c + \left(N_q + 1\right) + \left(N_p + 1\right) + \max(d_{11}, d_{12}, d_{21}, d_{22}) - 1 \\ &= L_c + N_q + N_p + d_{\max} + 1, \end{aligned} \tag{36}$$

where $d_{\max} = \max(d_{11}, d_{12}, d_{21}, d_{22})$.

3.3. Computational Complexity Analysis. Now we discuss the computational complexity of the three methods (the least-squares method, the single-filter structure method, and the CAPZ method) from two aspects: crosstalk cancellation filter estimation and implementation. For the convenience of comparison, Table 1 lists some parameters for three methods, respectively, where the column "Inverse filter" denotes the filter resulted from matrix inversion (referring to (9), (18), and (34)), the column "Matrix size" denotes the size of the matrix being inverted. It should be noted that the term "inverse filter" is different from the term "crosstalk cancellation filter."

Table 1: Parameters for the three methods: the least-squares method, the single-filter structure method, and the CAPZ method.

Method	Inverse filter	Matrix size	Crosstalk cancellation filter length
Least-squares	h	$Size(G) = 2L_1 \times 2L_h$	$L_{h1} = L_h$
Single-filter structure	t	$Size(Q) = L_2 \times L_t$	$L_{h2} = L_t + L_g - 1$
CAPZ	c	$Size(B) = L_3 \times L_c$	$L_{h3} = L_c + N_p + N_p + d_{\max} + 1$

Table 2: Computational complexity of crosstalk cancellation filter estimation for the three methods: the least-squares method, the single-filter structure method, and the CAPZ method.

Method	Computation cost (in multiplications)
Least-squares	$8(O(L_{\text{inv}}^3) + 2L_{\text{inv}}^2 L_1)$
Single-filter structure	$O(L_{\text{inv}}^3) + 2L_{\text{inv}}^2 L_2$
CAPZ	$O(L_{\text{inv}}^3) + 2L_{\text{inv}}^2 L_3$

3.3.1. Computational Complexity of Crosstalk Cancellation Filter Estimation. From (9), (12), and (30), it is found that estimating the inverse filters h, t, and c consumes the major computation of crosstalk cancellation filter estimation. Thus only the computation of calculating the inverse filters is considered. Generally, the computational complexity of inverting a matrix of size $N \times N$ is $O(N^3)$, without taking advantage of matrix symmetry. The computation of estimating the inverse filters h, t, and c is closely related to the size of the matrix G, Q, and B, respectively. Supposing that the inverse filter lengths in the three methods are equal, that is, $L_h = L_t = L_b = L_{\text{inv}}$, we summarize the computational complexity in Table 2 for the three methods (referring to (9), (18), and (34)). The computational complexity is calculated in terms of multiplication. For example, when the size of G is $2L_1 \times 2L_h$, the number of calculations involved in matrix multiplication is $16L_h^2 L_1$, and matrix inversion is $O((2L_h)^3)$ (referring to (9), (10), and Table 1). Thus, the computation cost of the least-squares method is $8(O(L_h^3) + 2L_h^2 L_1)$, as listed in Table 2. The computation cost of the other two methods can be obtained in a similar way.

For the convenience of comparison, we rewrite the parameters L_1, L_2, and L_3 from Table 1 in an approximated form as

$$\begin{aligned} L_1 &= L_h + L_g - 1 \approx L_{\text{inv}} + L_g, \\ L_2 &= L_t + L_q - 1 = L_t + 2L_g - 2 \approx L_{\text{inv}} + 2L_g, \\ L_3 &= L_c + L_b - 1 = L_c + 2N_q + \Delta \approx L_{\text{inv}} + 2N_q. \end{aligned} \tag{37}$$

Generally, $L_g \gg N_q$ holds for a CAPZ model. Thus we have

$$L_2 > L_1 > L_3. \tag{38}$$

From Table 2, the computational complexity of the least-squares method is much higher than the other two methods (almost 8 times), while the computation of the single-filter structure method is a little higher than the proposed CAPZ method.

3.3.2. Computational Complexity of Crosstalk Cancellation Filter Implementation. The computational complexity of crosstalk cancellation implementation is proportional to the crosstalk cancellation filter length, as listed in Table 1. Since $L_g > N_p + N_q + d_{\max}$ holds for the CAPZ model, we have

$$L_{h1} < L_{h3} < L_{h2}, \tag{39}$$

with the assumption of $L_h = L_t = L_b$.

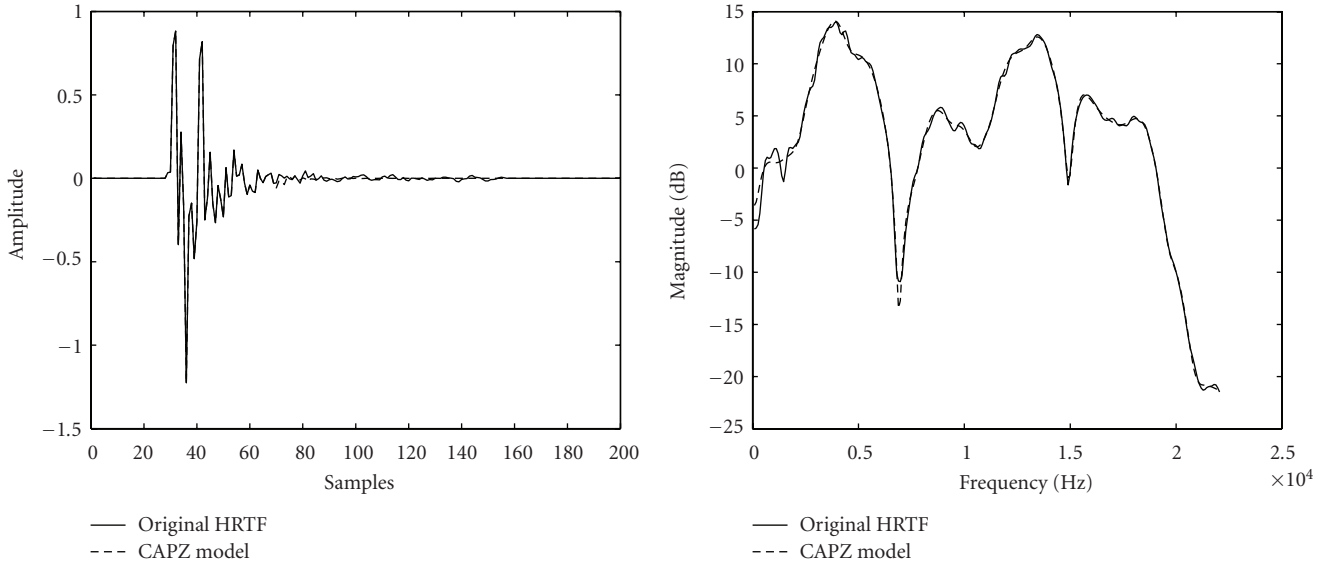
The least-squares method has the lowest computational complexity in crosstalk cancellation filter implementation, while the single-filter structure method has the highest one.

In summary, although the least-squares method has the lowest computational cost in filter implementation, its complexity in filter estimation is much higher than the other two. On the other hand, the CAPZ method has the lowest complexity in filter estimation, and ranks second in terms of the complexity of filter implementation. In a global view of both measures, the CAPZ method is the most effective among the three ones. Later, the performance comparison of the three methods will be carried out in Section 4.3 under the same assumption with $L_h = L_t = L_b = L_{\text{inv}}$.

4. Performance Evaluation

The acoustic transfer function can be estimated based on the positions of loudspeakers and ears. Head-related transfer functions (HRTF) provide a measure of the transfer path of a sound from some point in space to the ear canal. This paper assumes that the acoustic transfer function can be represented by HRTF in anechoic conditions. The HRTFs used in our experiments are from the extensive set of HRTFs measured at the CIPIC Interface Laboratory, University of California [29]. The database is composed of HRTFs for 45 subjects, and each subject contains 1250 HRTFs measured at 25 different azimuths and 50 different elevations. The HRTF is 200 taps long with a sampling rate of 44.1 kHz. In the experiment, the HRTFs are modeled as CAPZ models first, then the performance of the proposed crosstalk cancellation method is evaluated in two cases for loudspeakers placement: symmetric and asymmetric cases.

4.1. Experiments on CAPZ Modeling. For subject "003", the HRTFs from all 1250 positions are approximated with CAPZ models. Before modeling, the initial delay of each HRIR is recorded and removed. The common pole number is set empirically as $N_p = 20$, and the zero number $N_q = 40$. The original and modeled impulse responses and magnitude responses of the right ear HRTF at elevation 0°, azimuth 30° are shown in Figures 2(a) and 2(b), respectively. It can be seen from these figures that only small distortions can be noticed between the original and modeled HRTFs. Similar results may be observed at other HRTF positions.



(a) Impulse responses of the original and modeled HRTFs

(b) Magnitude responses of the original and modeled HRTFs

Figure 2: Comparison of the original and modeled right ear HRTF at elevation 0°, azimuth 30°.

4.2. Performance Metrics. Two performance measures are used: the signal-to-crosstalk ratio (SCR) and the signal-to-distortion ratio (SDR) [8]. Regarding to (6), the ideal crosstalk cancellation result should be

$$\mathrm{GH} = U = \begin{bmatrix} u_1 & O \\ O & u_2 \end{bmatrix}. \tag{40}$$

Since G is generally nonminimum-phase, the actual crosstalk cancellation result is

$$\mathrm{GH} = F = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}. \tag{41}$$

The signal-to-crosstalk ratio at two ears would be

$$\mathrm{SCR}_1 = \frac{f_{11}^T f_{11}}{f_{12}^T f_{12}}, \quad \mathrm{SCR}_2 = \frac{f_{22}^T f_{22}}{f_{21}^T f_{21}}, \tag{42}$$

and the average signal-to-crosstalk ratio is given by $\mathrm{SCR} = (\mathrm{SCR}_1 + \mathrm{SCR}_2)/2$.

And the signal-to-distortion ratio at two ears is determined by

$$\begin{aligned} \mathrm{SDR}_1 &= \frac{1}{(f_{11} - u_1)^T (f_{11} - u_1)}, \\ \mathrm{SDR}_2 &= \frac{1}{(f_{22} - u_2)^T (f_{22} - u_2)}, \end{aligned} \tag{43}$$

and the average signal-to-distortion ratio is $\mathrm{SDR} = (\mathrm{SDR}_1 + \mathrm{SDR}_2)/2$.

According to the definitions above, the signal-to-crosstalk ratio measures the crosstalk suppression performance, and signal-to-distortion ratio measures the signal reproduction performance.

4.3. Performance Evaluation in Symmetric Cases. In this experiment, the loudspeakers are placed in symmetric positions. Three crosstalk cancellation methods are compared: the least-squares method, the single-filter structure method, and the proposed method based on CAPZ models. To be consistent with the assumption in computational complexity analysis in Section 3.3, the inverse filter lengths in the three methods are set equal, that is, $L_h = L_t = L_c$. A total of 63 crosstalk cancellation systems are designed at 7 different elevations uniformly spaced between 0° and 67.5° and 9 different azimuths uniformly spaced between 5° and 45°. For each crosstalk cancellation system, various inverse filter lengths ranging from 50 to 400 samples with an interval of 50 are tested. Generally, the crosstalk cancellation performance is not quite sensitive to the delay value; however, an optimal delay value is selected for each method separately so that they can be compared in a fair condition. Since the relationship between the crosstalk cancellation and the delay z^{-d} shows no evident regularity, we choose the delay value experimentally. For each experiment case, the optimal delay is selected experimentally from values ranging from 50 to 400 samples with an interval of 50, ensuring that the crosstalk cancellation algorithm performs best with this optimal delay. Table 3 lists the optimal delay for the three methods at various inverse filter lengths. The regularization parameter is set empirically as $\beta = 0.005$ throughout the experiment. The mean value of the performance metrics over all 63 crosstalk cancellation systems is calculated.

Figure 3 shows the mean signal-to-distortion ratio (SDR), respectively, for the three methods with various inverse filter lengths. The horizontal axis is the inverse filter length ranging from 50 to 400 samples. The vertical axis is the mean signal-to-distortion ratio. The SDR of the least-squares method is always 2-3 dB higher than the CAPZ method, and 3-5 dB higher than the single-filter structure method.

Table 3: Optimal delay d at various inverse filter lengths (in samples) for the three methods: the least-squares method (LS), the single-filter structure method (SF), and the CAPZ method.

Filter length	LS	SF	CAPZ
50	50	100	100
100	100	150	150
150	100	150	150
200	150	200	200
250	150	200	200
300	200	250	250
350	200	250	250
400	250	300	300

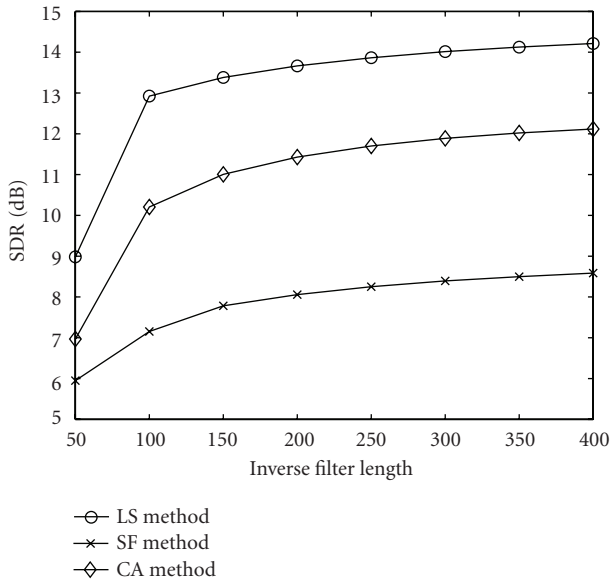


Figure 3: Mean signal-to-distortion ratio (SDR) at different inverse filter lengths for the three methods: the least-squares method (LS), the single-filter structure method (SF), and the CAPZ method.

Figure 4 shows the mean signal-to-crosstalk ratio (SCR), respectively, for the three methods with various inverse filter lengths. The horizontal axis is the inverse filter length ranging from 50 to 400 samples. The vertical axis is the mean signal-to-crosstalk ratio. Since the SCR of the SF method can be as high as 300 dB for all simulation cases, which is much higher than the levels of the other two methods (20–30 dB), its curve is left out of the picture. The SCR of the CAPZ is higher than the least-squares method. It can be seen from Figures 3 and 4 that the single-filter structure method yields the best SCR performance, while the least-squares method yields best SDR performance. On the other hand, for both SDR and SCR measures, the proposed CAPZ method yields performance that is superior to one of the reference methods, but inferior to the other reference. In a view of crosstalk cancellation, the performance of the CAPZ method is in the middle of the three methods. It can yield comparable crosstalk cancellation as the other two methods do.

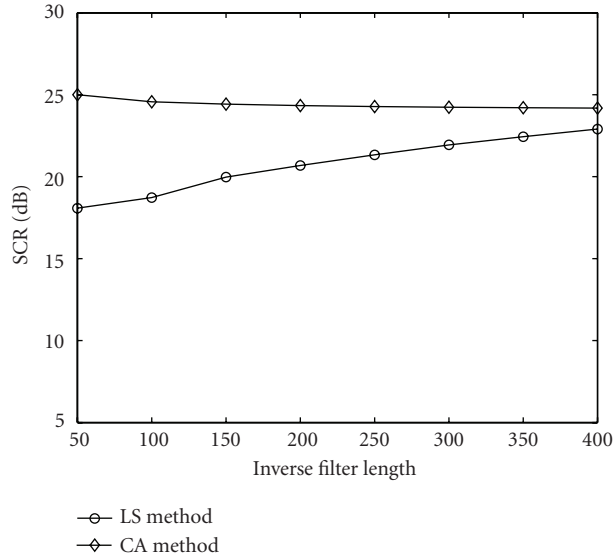


Figure 4: Mean signal-to-crosstalk ratio (SCR) at different inverse filter lengths for the three methods: the least-squares method (LS), the single-filter structure method (SF), and the CAPZ method. (Note that the curve of the SF method is not depicted in the picture, because its SCR values can be as high as 300 dB for all simulation cases.)

As discussed at the end of Section 2, with the off-diagonal items of the global transfer function (21) being zeros, the single-filter structure method can obtain nearly perfect crosstalk suppression. That is why the signal-to-crosstalk ratio (SCR) can be as high as 300 dB, which is implied in Figure 4. In practice, inevitable errors in the measurement process (nonideal HRTFs) result in degraded performance. To conduct a more realistic evaluation, we add random white noises with a signal-to-noise ratio of 30 dB to the HRTF measurement, and repeat the previous experiment. Although this is not a real non-ideal HRTF, the white noise may partly simulate errors and disturbances encountered during the measurement. This process is repeated five times, and then an average result is calculated. The mean signal-to-distortion ratio and signal-to-crosstalk ratio of the three methods are shown in Figures 5 and 6, respectively. The result is similar to the noise-free case: the performance of the three methods all decreases a little; especially, the SCR of the single-filter structure method reduce to about 26 dB.

From Figures 3–6, similar variation trends of the signal-to-distortion ratio (SDR) and signal-to-crosstalk ratio (SCR) may be observed for both noisy and noise-free cases. For all the three methods, the SDR performance increases with the inverse filter length L_{inv}, and the increase is small for $L_{\text{inv}} > 150$. The slow variation of SDR for large L_{inv} may be related to the least-squares matrix inversion process. When L_{inv} increases, the size of the matrices G, Q and B increases, the matrix inversion becomes difficult and more errors will be introduced. The error may cancel part of the benefit brought by a longer inverse filter. Thus the SDR increases slowly for large inverse filter length. With regard to the SCR performance, the least-squares method yields increasing SCR

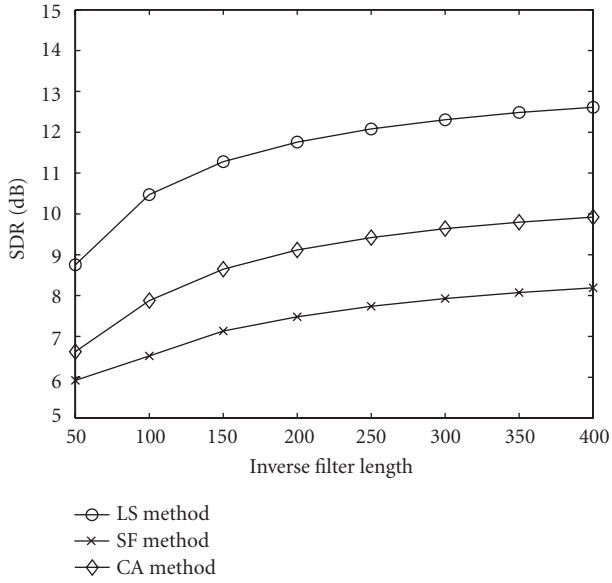


Figure 5: Mean signal-to-distortion ratio (SDR) at different inverse filter lengths for the three methods: the least-squares method (LS), the single-filter structure method (SF), and the CAPZ method (white noise added to HRTF).

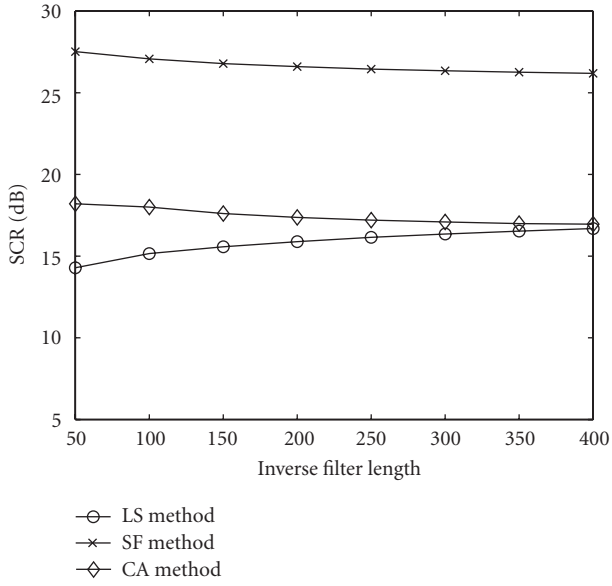


Figure 6: Mean signal-to-crosstalk ratio (SCR) at different inverse filter lengths for the three methods: the least-squares method (LS), the single-filter structure method (SF), and the CAPZ method (white noise added to HRTF).

with the increasing inverse filter length, while the single-filter structure method and the CAPZ method yield almost constant SCR with the increasing inverse filter length. Since the off-diagonal items of (21) are always zeros regardless of the value of $T(z)$, the SCR of the single-filter structure method is little affected by the inverse filter length. Likewise, the CAPZ method shows similar trend as the single-filter structure method does. In Figure 6, a slow decrease is also noticed for the curves of the CAPZ method and the single-filter structure method, which may be caused by the noise added to the acoustic transfer functions.

Table 4: Mean crosstalk cancellation performance in the symmetric case for the three methods when the inverse filter length equals 150.

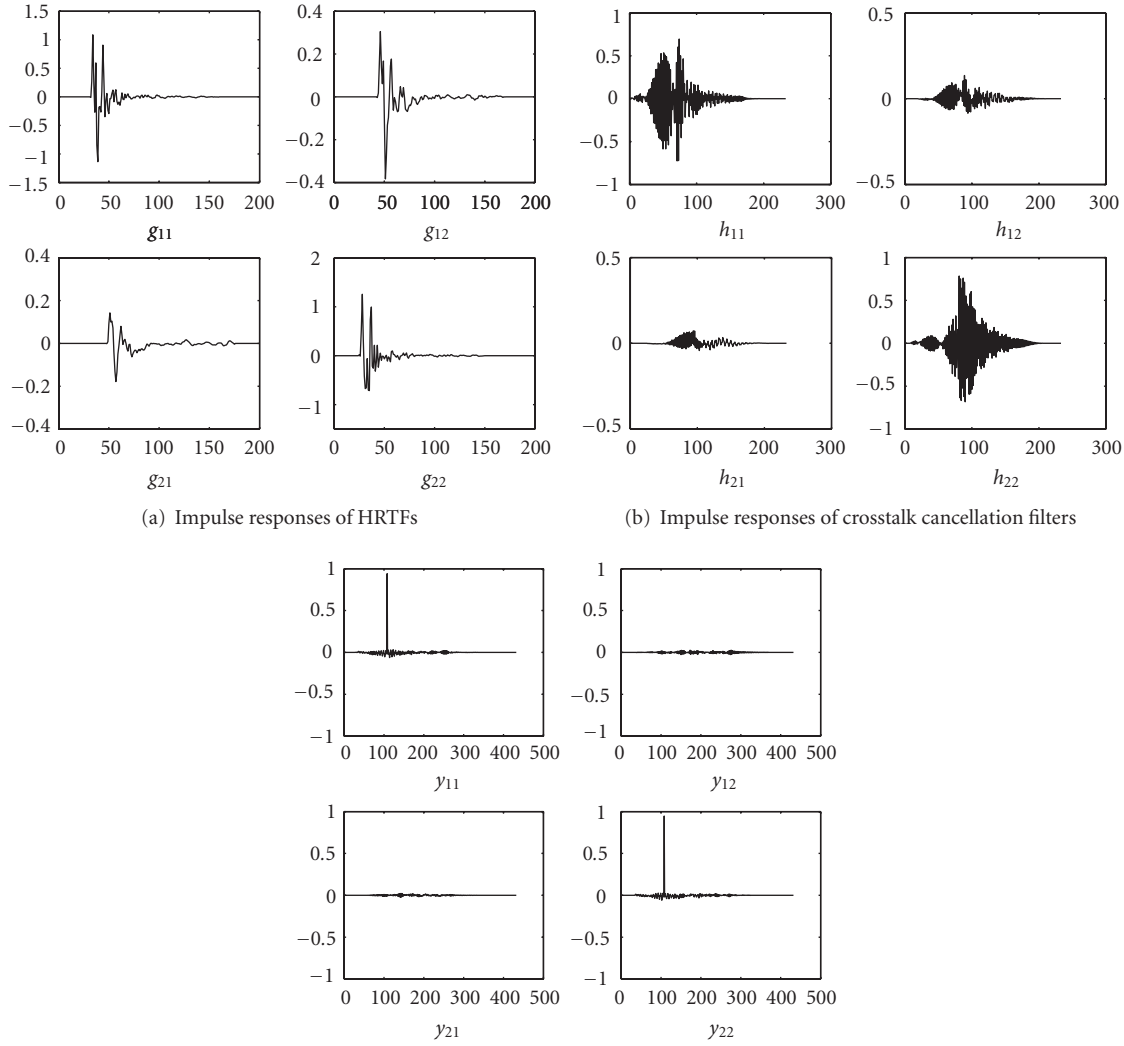
Method	SDR(dB)	SCR(dB)	Crosstalk cancellation filter length
Least-squares	11.2	15.6	150
Single-filter structure	7.1	26.8	349
CAPZ	8.6	17.6	233

Table 5: Crosstalk cancellation performance in the asymmetric case for the three methods when the inverse filter length equals 150.

Method	SDR(dB)	SCR(dB)
Least-squares	14.7	18.9
Single-filter structure	10.2	27.7
CAPZ	12.0	19.1

In summary, the proposed CAPZ method yields similar crosstalk cancellation performance as the other two methods do, meanwhile it is more computationally efficient. In a global view of both crosstalk cancellation and computational complexity, the proposed method is superior to the other two methods. Taking both performance and computation into consideration, we set the inverse filter length at 150. When white noises with a signal-to-noise ratio of 30 dB is added to HRTF, the performance of the three methods are listed in Table 4. The result in Table 4 also verifies the conclusion above.

4.4. Performance Evaluation in Asymmetric Cases. In this experiment, the stereo loudspeakers are placed in asymmetric positions, with the left and right loudspeakers at 30° and 60°, respectively, equidistant from the listener. Although this is not a common audio system, the crosstalk canceller can reproduce the desired sound field around the listener. The inverse filter length is set at 150, the regularization parameter is set at $\beta = 0.005$, the filter delay d is chosen from Table 3, white noise with a signal-to-noise ratio of 30 dB is added to the HRTF measurement. The performance of the three methods is shown in Table 5. Comparing Table 4 with Table 5, it can be seen that the performance of the three methods in the asymmetric cases is similar to that in the symmetric case. To give the readers a better understanding of the principle of crosstalk cancellation, Figure 7 depicts the impulse responses of the crosstalk cancellation system by the CAPZ method. The impulse responses of the HRTFs of 200 taps are shown in Figure 7(a), the four crosstalk cancellation filters designed by the CAPZ method are shown in Figure 7(b), and the result impulse responses after crosstalk cancellation are shown in Figure 7(c). Clearly, a good crosstalk cancellation can be obtained.



(a) Impulse responses of HRTFs

(b) Impulse responses of crosstalk cancellation filters

(c) Resulted impulse responses after crosstalk cancellation

Figure 7: Impulse responses of crosstalk cancellation in the asymmetric case.

5. Conclusion

This paper investigates crosstalk cancellation for authentic binaural reproduction of stereo sounds over two loudspeakers. Since the crosstalk cancellation filter has to be updated according to the head position in real time, the computational efficiency of the crosstalk cancellation algorithm is crucial for practical applications. To reduce the computational cost, this paper presents a novel crosstalk cancellation system based on common-acoustical pole/zero (CAPZ) models. The acoustic transfer paths from loudspeakers to ears are approximated with CAPZ models, then the crosstalk cancellation filter is designed based on the CAPZ model. Since the CAPZ model has advantages in storage and computation, the proposed method is more efficient than conventional ones. Simulation results demonstrate that the proposed method can reduce the computational complexity greatly with comparable crosstalk cancellation performance with respect to conventional methods.

The experiment in this paper is conducted in anechoic conditions. However, with promising results in anechoic environments, the proposed method can be extended to realistic situations. For example, in reverberation conditions, the acoustic transfer functions may also be approximated by the CAPZ model, and then crosstalk cancellation may be conducted in a similar way. However, due to large computational complexity and time-varying environments, this situation has not been specially addressed. Our further research will focus on this practical problem.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (60772161, 60372082) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (200801410015). This work is also supported by NRC-MOE Research and Postdoctoral Fellowship

Program from Ministry of Education of China and National Research Council of Canada.The authors gratefully acknowledge stimulating discussions with Dr. Heping Ding and Dr. Michael R. Stinson from Institute for Microstructural Sciences, National Research Council Canada.

References

[1] D. R. Begault, *3D Sound for Virtual Reality and Multimedia*, Academic Press, London, UK, 1st edition, 1994.

[2] A. W. Bronkhorst, "Localization of real and virtual sound sources," *Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2542–2553, 1995.

[3] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995.

[4] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2589–2598, 2006.

[5] B. S. Atal and M. R. Schroeder, "Apparent sound source translator," US Patent no. 3,236,949, 1966.

[6] B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," *Journal of the AudioEngineering Society*, vol. 9, no. 2, pp. 148–151, 1961.

[7] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, 1998.

[8] Y. Huang, J. Benesty, and J. Chen, "On crosstalk cancellation and equalization with multiple loudspeakers for 3-D sound reproduction," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 649–652, 2007.

[9] J. Garas, *Adaptive 3D Sound Systems*, Kluwer Academic Publishers, Norwell, Mass, USA, 2000.

[10] A. Mouchtaris, P. Reveliotis, and C. Kyriakakis, "Inverse filter design for immersive audio rendering over loudspeakers," *IEEE Transactions on Multimedia*, vol. 2, no. 2, pp. 77–87, 2000.

[11] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1621–1632, 1992.

[12] A. Gonzalez and J. J. Lopez, "Time domain recursive deconvolution in sound reproduction," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 833–836, June 2000.

[13] S. M. Kuo and G. H. Canfield, "Dual-channel audio equalization and cross-talk cancellation for 3-D sound reproduction," *IEEE Transactions on Consumer Electronics*, vol. 43, no. 4, pp. 1189–1196, 1997.

[14] C. Kyriakakis, "Fundamental and Technological Limitations of Immersive Audio Systems," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 941–951, 1998.

[15] M. R. Bai and C.-C. Lee, "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," *Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 1976–1989, 2006.

[16] T. Lentz, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," *Journal of the Audio Engineering Society*, vol. 54, no. 4, pp. 283–294, 2006.

[17] D. B. Ward and G. W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," *IEEE Signal Processing Letters*, vol. 6, no. 5, pp. 106–108, 1999.

[18] T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," *Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 2786–2797, 2002.

[19] M. R. Bai, C.-W. Tung, and C.-C. Lee, "Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm," *Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 2802–2813, 2005.

[20] J. Yang, W.-S. Gan, and S.-E. Tan, "Improved sound separation using three loudspeakers," *Acoustic Research Letters Online*, vol. 4, pp. 47–52, 2003.

[21] Y. Kim, O. Deille, and P. A. Nelson, "Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners," *Journal of Sound and Vibration*, vol. 297, no. 1-2, pp. 251–266, 2006.

[22] M. Kallinger and A. Mertins, "A spatially robust least squares crosstalk canceller," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, pp. 177–180, April 2007.

[23] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 320–328, 1994.

[24] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 188–195, 1999.

[25] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.

[26] S.-M. Kim and S. Wang, "A Wiener filter approach to the binaural reproduction of stereo sound," *Journal of the Acoustical Society of America*, vol. 114, no. 6, pp. 3179–3188, 2003.

[27] L. Wang, F. Yin, and Z. Chen, "HRTF compression via principal components analysis and vector quantization," *IEICE Electronics Express*, vol. 5, no. 9, pp. 321–325, 2008.

[28] D. W. Grantham, J. A. Willhite, K. D. Frampton, and D. H. Ashmead, "Reduced order modeling of head related impulse responses for virtual acoustic displays," *Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 3116–3125, 2005.

[29] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 99–102, October 2001.